Response to Reviewers " Nonexistence of unbiased test for high-dimensional linear model and a Bayesian-motivated test "

Rui Wang[1] and Xingzhong Xu[*1,2]

[1] School of Mathematics and Statistics, Beijing Institute of Technology, Beijing 100081,China

[2] Beijing Key Laboratory on MCAACI, Beijing Institute of Technology, Beijing 100081,China

Monday 30th September, 2019

We thank both reviewer for their helpful comments and critiques. We have carefully considered the comments and made corresponding changes to the paper. Our responds are as follows.

1 Response to reviewer 1

 1. The manuscript should provide comparison with the literature that exploits sparsity because under the null hypothesis the model parameter is sparse (with at most q non-zero entries). For example, Zhang and Cheng (2017) studies the same problem of testing (high-dimensional components) of linear regressions. Apart from assuming random designs, this literature attempts to detect deviations in l_∞-norm, while the current manuscript seems to aim at detecting deviations in l_2-norm. In my opinion, more discussion on this literature would be helpful so that the reader would get a fair impression on what has been done and what assumptions different methods require. For example, here symmetry of error distribution is needed (Assumption 1), which is not usually required in the literature on high-dimensional inference.

Answer: Following the reviewer's suggestion, we have provided the simulation results of the desparsifying lasso test of Zhang and Cheng (2017) (abbreviated as DL) and add the following comment to clarify the application regime of the proposed test.

> In the sparse $\boldsymbol{\beta}_b$ setting of Model II and Model III, the proposed test is not as powerful as the DL test. In fact, as an extreme value type test, the DL test is designed for

---
[*]Corresponding author

Email address: xuxz@bit.edu.cn

sparse linear models. In contrast, the proposed test test is a quadratic form which offers higher power when the signal is dense. On the basis of the simulations, we can recommend using the proposed test when there is no prior knowledge of sparsity.

As the reviewer pointed out, the symmetry of error distribution is not usually required in the literature on high-dimensional inference. However, this condition is a key assumption in the asymptotic theory of quadratic forms and is also assumed by Bai et al. (2018). It will be a highly nontrivial task to relax this assumption. We leave it for possible future work. We add the following comment in Section 5:

> Our Theorem 2 gives a general approximation for the distributions of quadratic forms. A key assumption in Theorem 2 is that the distribution of ξ_1 is symmetric about 0.. This condition is also assumed by Bai et al. (2018) in the study of central limit theorem of quadratic form. However, this condition is not usually required in the literature on high-dimensional inference. We think it will be an interesting and useful work to relax the symmetric condition in Theorem 2.

2. Theorem 4 provides the power analysis, but its assumption in Equation (8) needs more discussion. For example, at the end of Section 3, the manuscript uses the factor model to illustrate the power of the proposed test over Goeman et al. (2006). However, this discussion is based on Theorem 4 and my rough calculation suggests that Equation (8), which is required by Theorem 4, does not seem to hold for factor models. Let me outline some details.

(a) **Supppose that $q = 0$ (so $\tilde{\mathbf{U}}_a = \mathbf{I}_n$) and $\mathbf{X}_b = \begin{pmatrix} W_1^\top \\ \vdots \\ W_n^\top \end{pmatrix}$ with $W_i = BF_i + u_i$, where F_i is a scalar from i.i.d. $\mathcal{N}(0,1)$, u_i is from $\mathcal{N}(0, \mathbf{I}_p)$ and B is from $\mathcal{N}(0, \mathbf{I}_p)$. Assume F_i, u_i and B are independent.**

(b) **Since $\mathrm{E}(\gamma_I^{2k}) \geq \mathrm{Var}(\gamma_I^k)$, a necessary condition for Equation (8) is**

$$\frac{(\gamma_1^k - \mathrm{E}_I(\gamma_I^k))^2}{n\, \mathrm{E}_I(\gamma_I^{2k})} \to 0,$$

where E_I denotes expectation taken over randomness of I.

(c) **Based on the computation above Theorem 4 (i.e. $\mathrm{E}_I(\gamma_I^k) = \mathrm{tr}(\mathbf{X}_b\mathbf{X}_b^\top)^k/n$), we know that $\mathrm{E}_I(\gamma_I) = n^{-1}\sum_{i=1}^n W_i^\top W_i$ and**

$$\mathrm{E}_I(\gamma_I^2) = n^{-1}\sum_{i=1}^n (W_i^\top W_i)^2 + n^{-1}\sum_{i=1}^n \sum_{j=1, j\neq i}^n (W_i^\top W_j)^2.$$

(d) **For $k = 1$, we use the assumptions in (a) to determined the size of γ_1, $\mathrm{E}_I(\gamma_I)$ and $\mathrm{E}_I(\gamma_I^2)$. Simple computations would give use $\mathrm{E}\, W_i^\top W_i = 2p$, $\mathrm{E}(W_i^\top W_i)^2 = 6p^2 + 10p$ and**

2

$\mathrm{E}(W_i^\top W_j)^2 = p^2 + 5p$ **for** $i \neq j$. **Thus,** $\gamma_1 = O_p(np)$, $\mathrm{E}_I(\gamma_I) = O_p(p)$ **and** $\mathrm{E}_I(\gamma_I^2) = O_p(np^2)$. **Thus,** $\frac{(\gamma_I^k - \mathrm{E}_i(\gamma_I^k))^2}{n\,\mathrm{E}_I(\gamma_I^{2k})}$ **with** $k = 1$ **seem to be of the order** $O_P(1)$ **instead of** $o_P(1)$.

(e) **Overall, I think discussions on why Equation (8) should hold is needed, at least for the examples used at the end of Section 4 (i.i.d. rows and factor models).**

Answer:

3. I think the power comparison with Goeman et al. (2006) should be more formal. Currently, there is a comparison on when $\mathrm{Cov}(-\gamma_I^{-1}, \gamma_I w_I^2)$ **and** $\mathrm{Cov}(\gamma_I, \gamma_I \omega_I^2)$ **are positive. However, in the two expressions for asymptotic power on page 12, this is not the only difference. For example, there is also** $\mathrm{Var}(\gamma_I)$ **versus** $\mathrm{Var}(\gamma_I^{-1})$. **Therefore, in the end, which one is more powerful is not completely obvious to me. Even in the simple case of** $w_i = 0$ **for** $i = 1, ..., nq$, **I am not certain which test has better asymptotic power. I think a formal result comparing the power of the two test would be nice even under restricted assumptions.**

Answer:

2 Response to reviewer 2

1. The title should be changed. Nonexistence of unbiased test is not the main focus of this article, it is just the motivation of the paper. I suggest to change the title to "A Bayesian test for high dimensional linear regression models with fixed design matrix".

Answer: We agree that the original title is not good enough. We partly follow the suggestion of the reviewer and change the title to "A Bayesian motivated test for high dimensional linear regression models with fixed design matrix". We preserve the word "motivated" since Bayesian method is just the motivation of our test and our test is actually a frequentist test.

2. The Introduction part should be re-written. The main focus of this article is two-aspects. One is the Bayesian test, and the other is the fixed design matrix. The introduction part should clearly state about your contribution through the two aspects compared with other competitors.

Answer: We follow the advice of the reviewer and extend the introduction part.

3. The transition through Section 2 to Section 3 is not clear. When you considering tests with good average power. Why Bayesian methods are natural choices? Some discussions are needed. In addition, the non-existence of nontrivial unbiased test is a trivial result for high dimensional setting.

Answer: In frequentist point of view, Bayesian methods aim at minimize the average risk. See Lehmann's book. In the revised paper, we add some discussion at the end of Section 2.

4. Your proposed test is quite similar with that of the likelihood ratio test, some discussions for the differences are needed. In addition, since $\mathbf{X}_b^*$ **is of dimension** p,

when both n and p are large, the matrix $\mathbf{X}_b^* \mathbf{X}_b^{*\top}$ is not invertible. Does your method applicable for $n > p$?

Answer: We follow the suggestion of the reviewer. Our method is not applicable for $n > p$. If $n > p$, the prior distributions adopted by us will yield exactly the likelihood ratio test. This is an advantage of out test as it can be seen as a natural extention of the likelihood ratio test to the high dimensional setting. When $n > p$, the likelihood ratio test exists and should be used.

5. The differences between your test and that of Goeman et al. (2006) and Lan et al. (2014) need to be discussed carefully.

Answer: The first difference is the test statistic. We use the different prior strategy, and produce different statistics. The second difference is the way to determine the critical value. Our method has a wider application scope.

6. Throughout the article, I do not see any results related to fix design matrix, why the fix design is essential for your analysis? Is it useful for your theoretical results?

Answer: The result in Section 2 requires the fix design matrix. In fact, if the design matrix is random, there indeed exists tests with nontrivial power. See the papers on minimax test. For Bayesian methods, fixed design and random design will produce the same statistic. As for the theoretical results for our test. These results also hold for random design matrix provided the random design matrix satisfies our conditions.

3   List of major changes

References

Bai, Z., Pan, G., and Yin, Y. (2018). A central limit theorem for sums of functions of residuals in a high-dimensional regression model with an application to variance homoscedasticity test. *TEST*, 27(4):896–920.

Goeman, J. J., van de Geer, S. A., and van Houwelingen, H. C. (2006). Testing against a high dimensional alternative. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 68(3):477–493.

Zhang, X. and Cheng, G. (2017). Simultaneous inference for high-dimensional linear models. *Journal of the American Statistical Association*, 112(518):757–768.